Quality aspects in price indices and international comparisons

van Mulligen, P.H.

6. *Summary and conclusions*

The main topic of this thesis has been the quality issue in price comparisons through time and across space. In studying this topic, this thesis has tried to answer the following questions: first, what is the impact of differences in quality on intertemporal and international price comparisons? Second, what are the theoretical and practical advantages and drawbacks of the hedonic method compared to the matched model method to deal with this 'quality issue'?
Below I briefly summarize the main conclusions from the preceding chapters. I will provide an assessment of how quality issues will affect the future of price measurement practice and the use of hedonics, and discuss the implications for measurement practices in construction of time series price indices and currency conversion factors.

The quality issue: 'inside the sample bias' and 'outside the sample bias'

Starting with the first question, which is the topic of Chapter 2, it is clear that quality differences occur for many goods and services. In a dynamic economy, new products appear continuously, and nearly as many disappear. Likewise existing products are continuously improved, and often provided at lower nominal prices given the improvement in quality characteristics. For accountants of economic activity, these changes pose a large problem on how to measure price changes which are controlled for quality improvement. In the case of international comparisons, the heterogeneity issue is even stronger. Not only are differences between similar goods larger across countries, the overlap in output mixes and expenditure packages is also smaller than in a single-country intertemporal case.
The quality issue works in two ways, i.e., as an 'inside the sample bias' and 'outside the sample bias'. The outside the sample bias relates to the entry and exit of products in the sample. The problem of neglecting new and disappearing products leads to non-representativeness of the price index sample and creates a bias the price measurement. When trying to compare 'like with like' either over time or across countries, the incorporation of new goods in a price index at the moment of introduction is not possible. New goods do not have a counterpart in a previous period, so no price change for such a good can be measured for its first phase of existence. Moreover most statistical offices only renew their price samples every five or even ten years, so that relatively new goods are only picked up after they have been around for several years.[1] The often dramatic price changes that new goods witness in their first years of existence are therefore often not measured at all. Although new goods generally do not have large shares in total expenditure, a bias is introduced never-

theless. A similar problem occurs when varieties ('new models') of existing products are introduced. Since the standard method of price index calculation is comparing prices of nearly identical products only, new models are often neglected in a similar way.

Like new products, products that disappear from the market also pose a problem when the price sample is kept fixed for a long period. When a product stops being available, its price can no longer be observed and a hole appears in the price index. Statistical offices sometimes try to fill these holes by estimating what the price of this product would have been if it still existed, occasionally on the basis of consulting firms that produced those items.

The best solution to diminish the outside the sample bias is to frequently update the sample of the price index, including the prices and weights of new products as soon as they become available, and deleting disappeared ones from the index as quickly as possible.

The second kind of bias is analogously called the 'inside the sample bias'. This bias occurs when prices of non-identical products are matched. This is especially apparent when a new variety of an existing product replaces an old variety. In the search for a replacement item, the statistician often links the prices of the two varieties, making an implicit or an (ad hoc) explicit adjustment for differences in specifications of the products. Various adjustment procedures are discussed in Chapter 2, all of which lead to biased estimates of price changes. The direction of the bias depends on the size and direction of both quality change and nominal price change.

When taken together these two problems combined can have a significant effect on the measurement of price change. The Advisory Report to the Senate Finance Committee (Boskin *et al.*, 1996) estimated that the new goods and quality bias resulted in an annual upward bias of the CPI in the U.S. of 0.6 percentage points. Similar studies in other countries showed comparable results. It is difficult to assess the plausibility of these 'net' estimates, as the exact magnitude of this bias is determined by 'gross' biases in thousands of individual sub-indices and their weights in the final index. Such biases may be offsetting each other to an important extent, and these offsetting effects may change over time and across countries. Regardless of the bias in the final index, an improper index methodology can have a large effect in the case of individual goods and services. Computers provide a well researched and very relevant example, but the problem of wrong index numbers is not confined to this product alone. Indeed not all prices of goods and services are upwardly biased, and some prices (e.g., services in a non-market environment) may in fact show a significant downward bias when it comes to a careful adjustment of quality characteristics (Triplett, forthcoming).

What has been said on price index numbers is equally relevant in comparisons of prices across countries, which are carried out to derive currency conversion factors such as purchasing power parities or unit value ratios. In both the expenditure approach and the industry-of-origin approach, prices of products are matched on the assumption that they are identical across countries.

Here we also encounter the two biases mentioned above. In particular the outside the sample bias is big in international comparisons of prices, especially when comparisons are made from the producer perspective. Some products (like aircraft or automobiles) are simply not produced in every country. Especially specialization on the basis of comparative advantages strengthens outside the sample biases.[29] Also increased differentiation in the product value chain, in many cases within vertically integrated multinational firms, can significantly affect adequate price comparisons across countries. Furthermore data sources from different countries can sometimes not be matched, or information is not available because of confidentiality reasons. Together this leads to a large percentage of output or expenditure that is not matched in international price comparisons. Since there is not a natural way to link countries to each others as with periods of time, quality differences between countries cannot be assumed to be as gradual as quality changes over time. Hence the outside the sample bias cannot be easily remedied, not even with larger samples of goods and services.[31]

Because of the lack of gradual differences in quality of goods and services and the lack of an obvious scaling of countries, the inside the sample bias is also more severe in international comparisons than in price index numbers. In case of the expenditure approach, which uses specification pricing, it may still be feasible to match items which are closely similar across countries, but such matched items may not be the most representative for both countries' expenditure patterns. The industry-of-origin approach, which uses unit values, is based on average 'prices' of aggregates of goods. Not only quality differences between individual products are then important, but also the composition of the aggregates for which unit values are calculated. The importance of this output mix problem was pointed out in Chapter 5. For international comparisons of output and productivity at industry level, the expenditure approach does not suffice because expenditure price comparisons only cover final products or services. In order to cover output of intermediate products, which is a substantial part of total output, industry of origin comparisons remain necessary, and the need to address the inside the sample bias therefore remains high on the agenda.

The hedonic method and its uses

I now turn to the second question addressed in this thesis. As the matched model method is in most fields still the default option of price measurement, I will discuss here the relative merits and drawbacks of the hedonic method.

Based on the premise that buyers are interested in consuming the *characteristics* of finished goods and services rather than the goods and services themselves, the hedonic method applies regression techniques to make an explicit quality adjustment on price indices or ratios. Because actual consumer behaviour obviously not only relates to final products, but also or rather to their

characteristics, this premise makes the theoretical case for the hedonic method very strong. Especially if one is interested in cost-of-living-indices, which hold consumer utility intact, the hedonic hypothesis seems inevitable.

Nevertheless, there have been theoretical objections against the hedonic method. The expansion of applications of the hedonic method in the U.S. has given rise to the revival of an old debate that has become known as the 'resource cost – user value' debate. Although in a sense somewhat of a separate issue, this debate has major theoretical implications for the use of hedonics. The core of the debate is the conflict between those who think that only variables that reflect a change in production costs should be taken into account when measuring price changes, and those who think that ultimately the value that users attach to a product (or a certain characteristic) is what counts. Triplett (1983) pointed out that both views have their merits depending on the approach of the price measurement. But from an economic point of view, user value and resource cost must converge. First, producers who do not take user value into account in their production process will lose market share to competitors that provide products and services with higher user value. Second, if in a free market economy the same user value is delivered at higher production cost than a competitor there will ultimately be a loss of business. Therefore this thesis holds the view that price measures should reflect both user value and producer costs, and that characteristics used for the estimation of hedonic functions should also reflect both.

Apart from theoretical considerations, practical ones have to be assessed as well, given the fact that the construction of price indices is of a highly empirical nature. Obviously, the greatest strength of the hedonic method compared with the matched model method is that it allows the statistician or researcher to calculate a price index or price ratio when the number of matched items is low or even zero. The hedonic method is therefore much better suited to deal with strong inside and outside the sample biases than the matched model method.

But as pointed out in Chapter 3, the hedonic method also has several practical drawbacks. One which is raised most often, is the lack of sufficient data on individual items and their characteristics. However, as discussed, this argument is a fallacy. The data problem for the hedonic method is not any bigger than for the matched model method. Products can only be matched if the analyst is sufficiently certain that they are (nearly) identical, which requires the same information on the characteristics of the product available. In fact the application of the hedonic method by statistical offices may not be that more expensive or data intensive after all, since much information on characteristics of goods is collected anyway for the purpose of matching. With the availability of detailed data on prices including characteristics (e.g., scanner data) such information is increasingly available. Summing up, the 'data argument' cannot be used to dismiss the hedonic method in favour of the conventional matched model method.

Relevant practical problems that need to be addressed when applying the hedonic method were discussed in more detail in Chapter 3. The bottom line of the assessment in that chapter is that the correct specification of a hedonic model is the crux to the construction of quality-adjusted price indices and currency conversion factors. In practice this may prove hard, because the choice of which characteristics to include is not straightforward. Furthermore, many 'desired' characteristics may prove unobservable or not quantifiable. The difficulty of specifying an adequate hedonic function will likely be the biggest obstacle to the practical application of the hedonic method.

This thesis includes two original empirical applications of the hedonic method. In Chapter 4, the hedonic method is applied on advertisement and scanner data for computers in the Netherlands. Price indices declined somewhat more than matched model indices, confirming the hypothesis that an index of a product with rapidly changing characteristics, like computers, is upwardly biased when unadjusted for quality changes. Although the difference between hedonic and matched model indices was not small, the effect of using an explicit quality adjustment by means of hedonics is already substantially reduced when using a chained index method rather than a fixed base. In the case of chaining, resampling takes care of much of the quality problem. Monthly chaining may not be necessary for every product, but yearly resampling of the price sample is a likely minimum. In this respect, in dealing with quality issues one has more to gain from frequent resampling than from the implementation of hedonic quality adjustment methods as such. Another important matter is the extensive use of weights for the individual items in the index. These conclusions were confirmed when the (high frequency) matched model and hedonic methods were applied on the database for the construction of the Dutch CPI for computers.

Although frequent resampling and the use of expenditure weights would therefore tremendously improve price indices, this can not always be realised in practice. Statistical offices have limited budgets for the collection of price data, and encounter special difficulties with the collection of expenditure shares. The hedonic method may provide a partial solution when the price sample is inadequate, especially when the number of matches is low and sampling is infrequent, but it does not provide a full and satisfactory remedy.

For international comparisons of producer prices of cars, which are studied in Chapter 5, there appears to be no overlap in items that can be matched. In different countries, different cars are produced, so applying the matched model method on an individual product basis is not possible. Using the hedonic method shows that average price ratios, unadjusted for quality differences, cause a severe understating of the real output value of the automobile industry for countries where on average larger and powerful cars are produced vis-à-vis the real output value in countries which produce smaller cars. However, the hedonic function for cars suffers from one major problem: the number of available characteristics is quite limited when set against the huge amount of price-determining characteristics of such complex products. In fact the available characteristics for cars (such as size and weight) are all proxies of the characteristics that actually matter.

Although the correct specification of hedonic functions for complex products like cars is very difficult and may be even impossible, we should not dismiss the results beforehand and stick with the unadjusted matched model results. Paying close attention to how we specify a hedonic function remains important, but one must accept the limited availability of data. By choosing the best possible specification, one should determine whether it is good enough, and judge the results by comparing them with matched model results. In some cases matching products may not be possible at all and one can then either has to rely on price relatives for related products in other industries or use the hedonic method, despite its limitations.

Because both the inside the sample and outside the sample biases are much bigger in international price comparisons than for price indices over time, the case for hedonic quality adjustments in the former case is in fact quite strong. However, the specification of a common hedonic function, as is possible for different periods (provided they are not too far apart), is much more difficult to apply across countries. The relation between prices and characteristics is determined by preferences and production processes, which are far less similar between countries than between relatively close periods within a single country. Only in very special cases with internationally standardised products the estimation of a common international hedonic function might be possible. Again computers may be a good example, as illustrated by Moch and Triplett (2002). For the data on cars used in Chapter 5, the relation between prices and characteristics differed too much between countries to employ a common hedonic function. The inadequacy of a common hedonic model in international price comparisons also precludes the use of the hedonic dummy method, which requires the pooling of data across countries.

The future of hedonics

Although the popularity of the hedonic method is still large under many scholars, it has come under recent attack from an American expert panel on price measurement (Schultze and Mackie, 2001). The Schultze Committee states that many econometric issues surrounding the hedonic method have still not adequately been resolved. Because the hedonic method has, in the Committee's view, not fully been developed, it states that the class of goods to which the hedonic method may be justifiably applied is narrow. However, much of the criticism and conservatism against the hedonic method expressed in this report is directed towards a variant of the hedonic method that is not advocated by many scholars, namely the dummy method.

The work by Schultze and Mackie ignores a lot of recent research on hedonic indices, as appears from a recent criticism by Diewert (2002). Many statistical offices have increased their use of the hedonic method compared with their previous methodologies. The views stated by Schultze and Mackie on hedonics may be shared by some statisticians and even economists, they are certainly not the *communis opinio* regarding the hedonic method.

Instead the debate on the hedonic method appears to be moving in favour of the strong advocates of hedonics, like Diewert (2002), Pakes (2002), Silver (2003) and Triplett (forthcoming).

Since this thesis puts the hedonic method as an appropriate tool for quality adjustment in price measurement, what is the best way forward in both statistical practice and academic application of hedonic price indices? As stated above, theoretical considerations are of importance when specifying a hedonic model. The choice of characteristics, and how to deal with proxies for unobservable characteristics are important issues to deal with for the correct specification of a hedonic model.

Provided such theoretical issues can be resolved, an important point is when to use the hedonic method as an instrument for quality adjustment. A distinction is made here between price indices across time and international comparisons. As noted above, for price index measurement, most can be gained by using a chained index principle with frequent resampling. The effect of hedonics in comparison with a the chain principle is modest. For price index numbers, hedonics therefore does need to be a high priority.

For international price comparisons, the need for the hedonic method is bigger. There is no international equivalent to a high frequency matched model index as an alternative method. Besides using price relatives from related products, as is common practice in current ICOP studies, hedonics is the only option especially in those fields where the overlap in production and expenditure of the same items are small. Collecting necessary data for international price comparisons is not an easy task, and using the hedonic method only makes it harder. In most cases it requires industry specific databases from, for example, multinational companies, that produce in more than one countries or marketing agencies or industry associations with offices in various countries. Without quality information, simply matching prices of products based on a similar functional use becomes an educated guess at best. This is especially true for complex products with many characteristics.

Finally, given the increased importance of the services sector in the economies of advanced countries, more attention needs to be paid to price measurement in services. Although this thesis focused on examining the quality issue in two types of (durable) goods, all of the above is equally relevant for services. As the measurement problems are huge, and the difficulty to distinguish prices, quantities and quality characteristics as set out in Chapter 1, the use of hedonic measurement may be even more relevant here. Services are often complex in terms of possessing a broad range of relevant characteristics, distinguishing them from other items. This is a problem for price measurement in general, be it hedonic or conventional measures. If we somehow can resolve these issues and enough information is available for matching prices, there is no reason why we cannot apply the hedonic method. As pointed out above, the only real drawback of the hedonic method compared with the matched model method is the difficulty of specifying an adequate hedonic function. As with physical goods, data issues are similar for both methods.

The measurement problems in the field of services are not easily overestimated. These are still huge, and preclude a rapid adoption of the hedonic method in services. But for physical goods, where the output concept is much clearer, the largest pay-off from using the hedonic method lies in international price comparisons.

Notes

[1] More frequent resampling is presently considered by many statistical offices, and has, for example, been implemented in the Dutch CPI in 2003.

[2] See, for example Dornbusch *et al.* (1977).

[3] Linking countries with the use of optimal spanning trees may resolve this problem to some extent. See Hill (1999).